Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000

Helene Feil*, William S. Feil*, Patrick Chain†‡, Frank Larimer§, Genevieve DiBartolo‡, Alex Copeland‡, Athanasios Lykidis‡, Stephen Trong‡, Matt Nolan‡, Eugene Goltsman‡, James Thiel‡, Stephanie Malfatti†‡, Joyce E. Loper¶, Alla Lapidus‡, John C. Detter‡, Miriam Land§, Paul M. Richardson‡, Nikos C. Kyrpides‡, Natalia Ivanova‡, and Steven E. Lindow*‖

*Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; ‡Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598; †Lawrence Livermore National Laboratory, Livermore, CA 94550; ¶U.S. Department of Agriculture, Agricultural Research Service, Corvallis, OR 97330; and §Oak Ridge National Laboratory, Oak Ridge, TN 37831

Contributed by Steven E. Lindow, June 14, 2005

The complete genomic sequence of *Pseudomonas syringae* pv. *syringae* B728a (*Pss* B728a) has been determined and is compared with that of *P. syringae* pv. *tomato* DC3000 (*Pst* DC3000). The two pathovars of this economically important species of plant pathogenic bacteria differ in host range and other interactions with plants, with *Pss* having a more pronounced epiphytic stage of growth and higher abiotic stress tolerance and *Pst* DC3000 having a more pronounced apoplastic growth habitat. The *Pss* B728a genome (6.1 Mb) contains a circular chromosome and no plasmid, whereas the *Pst* DC3000 genome is 6.5 mbp in size, composed of a circular chromosome and two plasmids. Although a high degree of similarity exists between the two sequenced Pseudomonads, 976 protein-encoding genes are unique to *Pss* B728a when compared with *Pst* DC3000, including large genomic islands likely to contribute to virulence and host specificity. Over 375 repetitive extragenic palindromic sequences unique to *Pss* B728a when compared with *Pst* DC3000 are widely distributed throughout the chromosome except in 14 genomic islands, which generally had lower GC content than the genome as a whole. Content of the genomic islands varies, with one containing a prophage and another the plasmid pKLC102 of *Pseudomonas aeruginosa* PAO1. Among the 976 genes of *Pss* B728a with no counterpart in *Pst* DC3000 are those encoding for syringopeptin, syringomycin, indole acetic acid biosynthesis, arginine degradation, and production of ice nuclei. The genomic comparison suggests that several unique genes for *Pss* B728a such as ectoine synthase, DNA repair, and antibiotic production may contribute to the epiphytic fitness and stress tolerance of this organism.

virulence genes | epiphyte | plant pathogen

Pseudomonas syringae, a member of the gamma subgroup of the Proteobacteria, is a widespread bacterial pathogen of many plant species. The species *P. syringae* is subdivided into ≈50 pathovars based on pathogenicity and host range. *P. syringae* is capable of producing a variety of different symptoms depending on the host species and site of infection. For example, it causes leaf-spot diseases that defoliate plants such as tomato, bean, and soybean, trunk cankers, and so-called "blast" diseases on fruit, nut, and ornamental species. Considerable variation occurs both between and within different pathovars of *P. syringae* (1). Because of its importance as a plant pathogen, it has been the subject of much research, especially of its epidemiology and virulence mechanisms (2). *P. syringae* pv. *syringae* (*Pss*) strain B728a is typical of most strains of this pathovar in that it exhibits a very pronounced epiphytic phase on plants. Such strains achieve and maintain large populations on healthy plants, where they are exposed to stressful conditions such as dryness and sunlight that are hostile to bacterial growth (2). Epiphytic *Pss* populations serve as inocula that can subsequently invade plants and initiate disease. *Pss* strains are distinct from many *P. syringae* strains, such as *P. syringae* pv. *tomato* (*Pst*) strain DC3000, that poorly colonize the exterior of plants; these strains may be considered "endophytes" based on their ability to multiply mostly within the plant (3). True epiphytes such as *Pss* B728a often reach surface populations of $>10^7$ cells per gram, whereas strains such as *Pst* DC3000 seldom exceed 10^5 cells per gram (2, 3). Thus, these strains might be considered to occupy different ends of the epiphytic/endophytic spectrum of plant colonization as described by Beattie and Lindow (4).

As a pathogen and an epiphyte, *Pss* B728a has evolved to exploit at least two distinct habitats: the leaf surface and apoplast. Because rapid changes in temperature, low water content, and incident solar radiation occur on leaf surfaces, it has been hypothesized that the epiphyte *Pss* B728a possesses more genes conferring environmental stress tolerance than the endophyte *Pst* DC3000 (4). *Pss* B728a also exhibits several traits such as ice nucleation activity and syringomycin (SR) production (2) that are lacking in some other strains of *P. syringae*, including *Pst* DC3000. As the most ice nucleation active bacterial species, *P. syringae* is responsible for inciting frost injury to frost-sensitive plants that can supercool and avoid damaging ice formation if not colonized by ice nucleation active bacteria (2, 4). We present here a genomic comparison between strains *Pss* B728a and *Pst* DC3000 of *P. syringae* pathovars as well as between these strains and *Pseudomonas aeruginosa* and *Pseudomonas putida*, two additional Pseudomonads recently sequenced. These genomic comparisons provide insights into the evolutionary history and diverse lifestyles of the pseudomonads, including their association with the environment, plant, or mammalian hosts.

Materials and Methods

Sequencing and Annotation Methods. *Pss* B728a was isolated from a snap bean leaflet in Wisconsin (5). The complete genome was sequenced at the Joint Genome Institute by using a combination of 3-kb and fosmid (40-kb) libraries. Library construction, sequencing, finishing, and automated annotation steps used were as described (6). Predicted coding sequences were subjected to manual analysis by using an Integrated Microbial Genomes annotation pipeline (http://img.jgi.doe.gov). Repetitive extragenic palindromes (REPs) were identified by using the sequence identified for palindromic repeats in *P. putida* (7) and the BLASTN search against the

Abbreviations: *Pss*, *P. syringae* pv. *syringae*; *Pst* DC3000, *P. syringae* pv. *tomato* DC3000; IS, insertion sequence; REPs, repetitive extragenic palindromic sequences; IAA, indole-3-acetic acid; AHL, acyl-homoserine lactone; SR, syringomycin.

Data deposition: The sequence and annotation reported in this paper have been deposited in the GenBank database (*Pss* B728a, accession no. CP000075).

‖To whom correspondence should be addressed. E-mail: icelab@socrates.berkeley.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0504930102

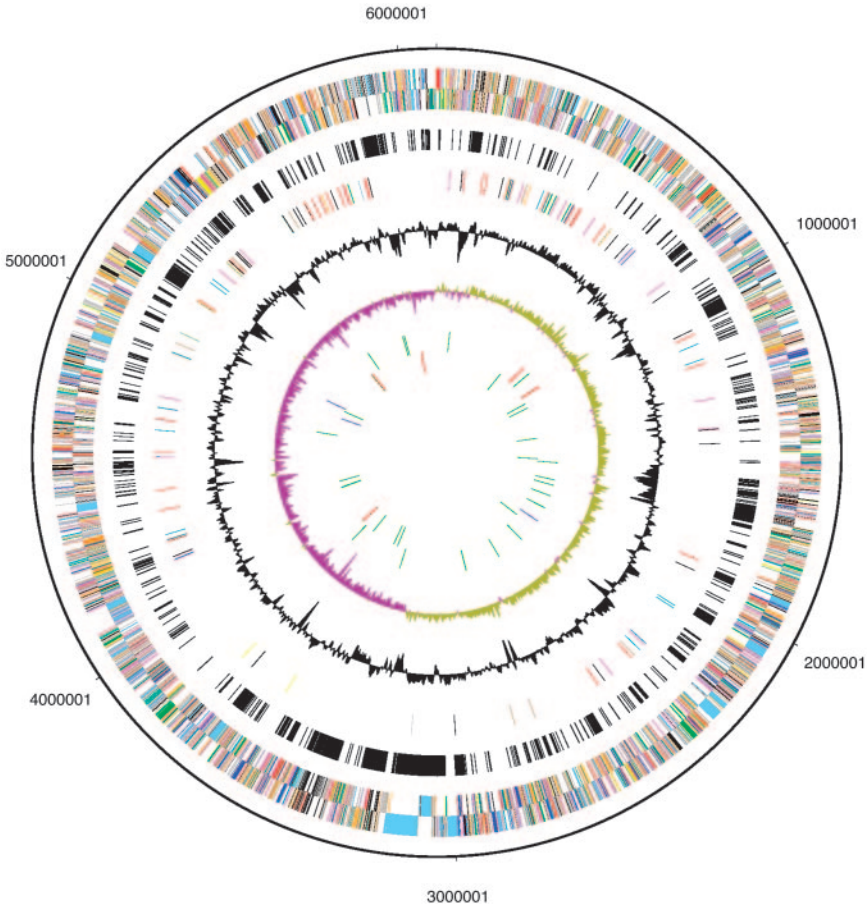

Fig. 1. Chromosome of *Pss* B728a with the outermost circle depicting ORFs on the positive strand and the second circle depicting ORFs on the negative strand. The ORFs are color-coded based on the major grouping of role categories as follows: hypothetical proteins (black); conserved hypothetical proteins (light gray); nucleotide metabolism and transport (light red); translation, ribosomal structure, and biogenesis (yellow); signal transduction (lavender); transcription (magenta); lipid metabolism (cyan); carbohydrate metabolism and transport (blue); coenzyme metabolism (pink); energy metabolism (green); DNA replication, recombination, and repair (red); general function prediction (brown); cellular processes (sky blue); amino acid metabolism and transport (orange); and structural RNA (pale green). The third circle represents the ORFs with no orthologs in Pst DC3000 ($E < 10^{-5}$. The fourth circle represents the palindromic REPs color-coded as follows: family 1 (dark gray), family 2 (red), family 3 (green), family 4 (blue), family 5 (cyan), family 6 (magenta), family 7 (yellow), family 8 (pale green), family 9 (sky blue), family 10 (orange), family 11 (brown), and family 12 (light gray). The fifth and sixth circles represent GC content and skew respectively. The seventh and eighth circles represent the tRNA (green), rRNA (red), and misc_RNA (blue).

Pss B728a genome sequence. Insertion sequence (IS) elements were identified by using the ISFinder database (www-is.biotoul.fr).

Comparative Genomics. Comparative analysis of *Pss* B728a and *Pst* DC3000 was performed by using a set of tools available in Integrated Microbial Genomes; BLASTP cutoff scores of $E < 10^{-2}$ and 20% identity were used to identify unique genes in *Pss* B728a and *Pst* DC3000, and cutoff scores of $E < 10^{-5}$ and 30% identity were used to find the homologous genes between *Pseudomonas* spp. LexA-binding sites were initially identified by using a consensus sequence for Gram-negative bacteria CTG-N_{10}-CAG, which was further refined based on the sequences found upstream of the known members of the LexA regulon.

Results and Discussion

Genome Features. The *Pss* B728a genome is composed of one circular chromosome of 6,093,698 bp (Fig. 1 and Fig. 4, which is published as supporting information on the PNAS web site) harboring 5,217 genes. The *Pss* B728a genome has 3,840 genes with predicted functions, 1,297 without known functions, and 80 RNA genes (Table 1). REP elements are 35-bp sequences of highly conserved inverted repeats with the potential of forming a stem–loop structure (7). At least 375 REPs occur in *Pss* B728a (Fig. 1). We identified 12 sequences characterizing REPs in *Pss* B728a (Table 3, which is published as supporting information on the PNAS web site). These repeated sequences are similar but distinct from the REPs found in *P. putida* (7).

Pss B728a shares 4,273 genes with *Pst* DC3000 and has 976 genes with no counterparts in *Pst* DC3000 (Table 2). *Pst* DC3000 has 5,608 protein-encoding genes, whereas *Pss* B728 has 5,090; therefore, *Pst* Dc3000 has 518 more protein-encoding genes than *Pss* B728a, most of which are transport and regulatory genes among the many duplicated genes in the *Pst* DC3000 genome (8). In addition, *Pst* DC3000 has 28 unique type III secretion system effectors (www-.Pseudomonas-syringae.org/pst_func_gen2.htm) and unique genes associated with coronatine biosynthesis. The remaining unique *Pst* DC3000 genes are categorized as encoding hypothetical or conserved hypothetical proteins. The vast majority of genes on the two plasmids of *Pst* DC3000 are not present in the *Pss* B728a genome. The genome of *Pss* B728a shares with *Pst* DC3000 most of the many genes whose functions are associated with transcriptional regula-

Table 1. Characteristics of the *Pss* B728a genome

Characteristic	No.	%
Genomic features		
Size, bp	6093698	
DNA coding sequence	5393016	88.50
% GC content		59.23
General statistics		
Total no. of genes	5,217	100.00
Protein-coding genes	5,137	98.47
RNA genes	80	1.53
rRNA genes	16	0.31
5S rRNA	6	0.12
16S rRNA	5	0.10
23S rRNA	5	0.10
tRNA genes	64	1.23
Genes with function prediction	3,840	73.61
Genes without function prediction	1,297	24.87
With similarity	1,255	24.06
Without similarity	42	0.81
Pseudogenes	47	0.90

Table 2. Global genomic comparison among four *Pseudomonas* spp.

Gene	*Pa*	*Pp*	*Pst*	*Pss*
Pa		2,927	3,441	2,969
Pp	3,201		3,017	2,624
Pst	3,449	2,859		976
Pss	3,444	2,816	1,335	
Pa + *Pp*			2,773	2,403
Pa + *Pst*		2,383		913
Pa + *Pss*		2,353	1,248	
Pp + *Pst*	2,938			884
Pp + *Pss*	2,952		1,260	
Pst + *Pss*	3,351	2,719		
Pa		2,365	2,117	2,122
Pp			2,591	2,534
Pst				4,273
Pss				

Top 10 lines, unique genes: Number of genes in the column organism without a homolog in the row organism(s). Bottom four lines, common genes: Number of genes in the column organism with a homolog in the row organism. *Pa* (*P. aeruginosa PAO1*), *Pp* (*P. putida KT2440*), *Pst* (*Pst* DC3000), *Pss* (*Pss* B728a). The calculations were based on a maximum *E*-value of 10^{-5} and minimum identity of 60%.

tion, signal transduction, transcription, gene replication, and membrane transport (Fig. 5, which is published as supporting information on the PNAS web site). The distribution of genes in various functional categories in *Pss* B728a and *Pst* DC3000 is similar. There are nine major rearrangements of the *Pss* B728a genome with respect to that of *Pst* DC3000 (Fig. 2). Because these two genomes each exhibit the expected G + C skews, we assume that the rearrangements are not recent, and that these two *P. syringae* pathovars diverged long ago.

Mobile genetic elements. The genome of *Pss* B728a has 16 IS elements, far fewer than the number found in *Pst* DC3000. *Pst* DC3000 has many more copies of each IS element than *Pss* B728a. The IS elements were found to be more frequent within the regions containing genes of unknown functions within the *Pst* DC3000 genome (9). Two IS elements (Psyr0106 and 0747) of *Pss* B728a are not present in *Pst* DC3000 (Table 4, which is published as supporting information on the PNAS web site).

Genomic islands. Many of the 976 unique genes of *Pss* B728a when compared with *Pst* DC3000 are found in 14 genomic islands (Table 5, which is published as supporting information on the PNAS web site). In particular, a genomic island for a prophage with many genes from phage LambdaSO from *Shewanella oneidensis* (10) occurs in *Pss* B728a. This region is flanked by tRNAs and therefore could have been horizontally acquired. Another genomic island inserted at tRNA-Lys-2 (position 1604394–1724175 on the *Pss* B728a genome) is related to the conjugative plasmid/genomic island pKLC102 of *P. aeruginosa* PAO1 (11). Similarly to pKLC102, it carries recombination and replication machinery and a *pil* operon encoding type IV thin sex pili. This island also encodes determinants of copper and arsenic resistance, none of which has homologs in Pst DC3000. As recently reported for *P. fluorescens* Pf-5 (12), REPs are largely absent from the genomic islands (Fig. 1). The lack of REPs suggests that these islands have a different evolutionary history from the rest of the genome.

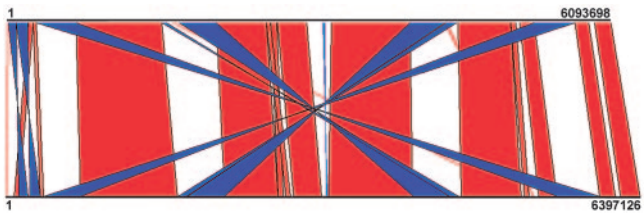

Fig. 2. Comparison of chromosome structure for *Pss* B728a vs. *Pst* DC3000. Axes represent the portions coded for in the order in which they occur in the chromosomes. Top axis, B728a; bottom axis, DC3000. Red represents colinear regions on the + strand of both genomes, and blue represents colinear regions that have been inverted in one of the two genomes. The display was generated by using ARTEMIS comparison tool (ACT) (www.sanger.ac.uk/software/artemis/ACT).

Lifestyle. ***Siderophores.*** Siderophores are low-molecular-weight high-affinity iron(III) chelators that are exported from the cell and that chelate iron in the extracellular milieu; the iron–siderophore complex is then transported into the cell via specific TonB-dependent receptor proteins, thereby providing iron for cellular functions. The fluorescent pseudomonads are characterized by the production of pyoverdines, a diverse class of siderophores containing a chromophore, which is responsible for the UV fluorescence typifying the group, linked to a small peptide of varied length and composition that is synthesized nonribosomally (13). In *Pss* B728a, as in *Pst* DC3000 (14), genes required for pyoverdine biosynthesis and uptake are present in a single gene cluster (Psyr1944–1961). Genes encoding for the biosynthesis and uptake of a second siderophore, probably yersinibactin, which are present in the genome of *Pst* DC3000 (13), are absent from *Pss* B728a. *Pss* B728a, however, has a second siderophore gene cluster composed of 14 genes with predicted biosynthetic, efflux, uptake, or regulatory functions. The six biosynthetic genes are closely related to genes for the biosynthesis of achromobactin (15), a citrate siderophore produced by *Pectobacterium chrysanthemi* (16) and apparently also by *Escherichia carotovora* pv. *atroceptica* (17). Linked to the biosynthetic genes are genes predicted to encode a TonB-dependent outer membrane receptor and four periplasmic or cytoplasmic membrane proteins also related to those functioning in achromobactin transport in *P. chrysanthemi* (15), as well as FecI/FecR homologs typically involved in regulation of siderophore-mediated iron acquisition (18). Multiple siderophores are required for full virulence of the plant pathogen *P. chrysanthemi* (15), and the capacity to draw from two distinct siderophore systems may also contribute to virulence or ecological fitness of *Pss* B728a.

UV resistance and tolerance to reactive oxygen species. To withstand UV-induced DNA damage, bacteria have developed a number of DNA repair mechanisms, including photoreactivation, base excision repair, nucleotide excision repair, and damage bypass. DNA polymerase V, encoded by *rulAB* in *P. syringae,* is an error-prone polymerase responsible for translesion synthesis on damaged DNA templates. This enzyme allows bacteria to cope with otherwise lethal effects of UV radiation, albeit at the price of an increased mutation rate.

Pss B728a is far less sensitive than *Pst* DC3000 to UV irradiation (19). Analysis of the genes involved in UV and reactive oxygen species (ROS) resistance in *Pst* DC3000 and *Pss* B728a (Table 6, which is published as supporting information on the PNAS web site) reveals that they share a number of pathways for detoxification of ROS and DNA repair, including photorepair, base excision repair, and nucleotide excision repair. Nevertheless, there are some differences between the two genomes that may explain the higher UV tolerance of *Pss* B728a. First, *Pss* B728a has five catalase isozymes, whereas *Pst* DC3000 has only three. Presumably, this would make the ROS quenching system of *Pss* B728a more efficient, which in turn could result in higher resistance to UVA irradiation. Second, *Pss* B728a has two copies of the *rulAB* operon. It has also an orphan *rulA* gene and remnants of two *rulAB* operons, one of them containing full-length *rulA* and truncated *rulB*, and the other containing two pseudogenes. The *rulAB* pseudogenes are highly similar to the full-length *rulAB* genes and appear to be the result of duplication and subsequent inactivation. In contrast, *Pst* DC3000 has fewer copies of *rulAB*, with a single *rulA* on plasmid A and one chromosomal *rulA* and *rulB*, which do not form an operon.

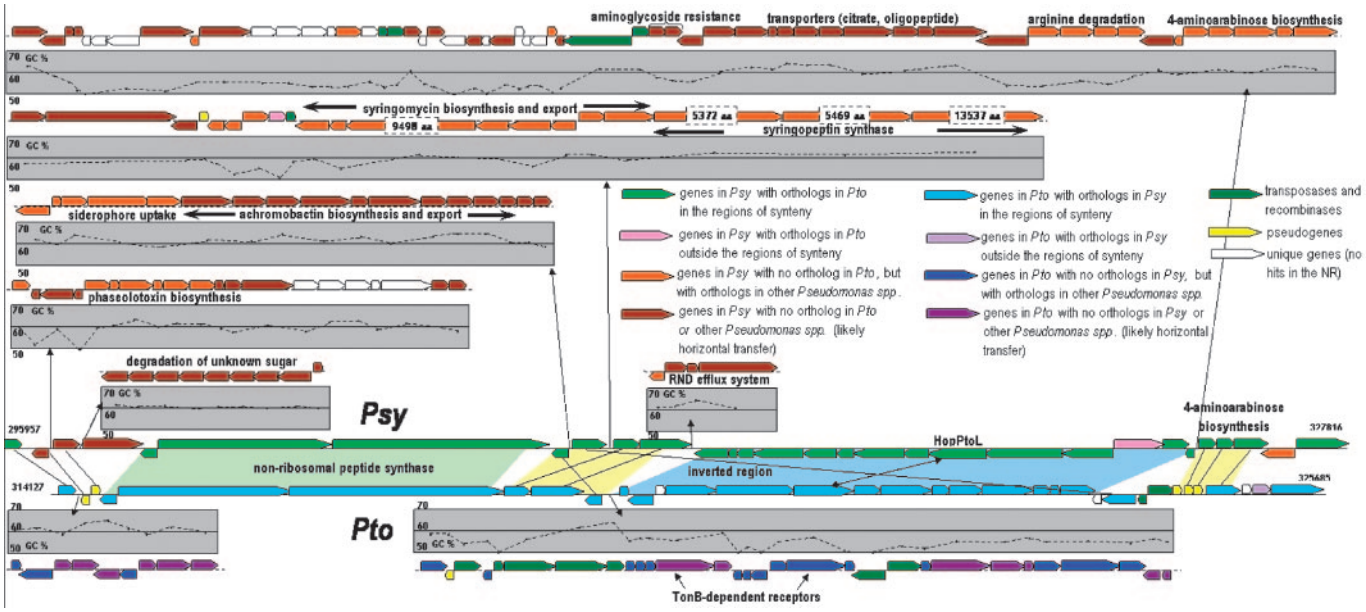

Fig. 3. Detailed view of the nonribosomal peptides and phytotoxins region of the *Pss* B728a genome. Comparison with the corresponding region for the *Pst* DC3000 is also shown.

The difference in UV resistance between *Pss* B728a and *Pst* DC3000 could also be due in part to differential expression of the multiple *rulAB* genes in *Pss* B728a. To be fully functional, *rulAB* must be expressed in a highly coordinated fashion in response to DNA damage. In most bacteria, synchronized induction of various DNA repair pathways is mediated via LexA, which represses genes by binding to a 16-mer consensus sequence CTG-N_{10}-CAG (20). Analysis of the potential LexA-binding sites in the *Pss* B728a genome shows that both *rulAB* operons are likely to be members of the LexA regulon (Table 7, which is published as supporting information on the PNAS web site). The finding of a LexA box with one mismatch upstream of the *rulAB* operon containing a full-length *rulA* gene and a *rulB* pseudogene suggests recent inactivation of this operon. No LexA boxes were detected in front of the other vestigial *rulAB* operon or the orphan *rulA* gene. The existence of multiple vestigial *rulAB* operons may reveal fluidity in mutagenic DNA repair systems in *Pss* B728a and suggests that its activity is regulated by changing the dosage of *rulAB* genes. The distribution of LexA-binding sites in *Pst* DC3000 tells a different story: whereas both copies of *rulA* belong to the LexA regulon, the *rulB* gene is devoid of a LexA box. Although the possibility exists that *rulB* is regulated by LexA indirectly via some unknown transcriptional regulator, it is likely that the *rulAB* system in *Pst* DC3000 is not fully functional due to insufficient expression of *rulB*.

Sequence variation within the *rulA* gene appears to have a significant impact upon UV tolerance of the host strain (21). A phylogenetic analysis of the *rulA* alleles found in different pseudomonads (Fig. 6, which is published as supporting information on the PNAS web site) shows that all *rulA* genes in *Pss* B728a and the chromosomal copy of *rulA* in *Pst* DC3000 (PSTO2121) are not clustered with the majority of *rulA* genes in other strains of *P. syringae*. The presence of the *rulAB* variants in highly UV-resistant *Pss* B728a calls for further experimental investigation.

Intra- and interspecies communication. Quorum sensing allows a population of a given bacterial species to synchronize behaviors that might be futile or detrimental unless done as a concerted action. Pseudomonads mediate quorum sensing by secretion of various types of autoinducers including *N*-acyl-homoserine lactones (AHLs), diketopiperazines, and furanosyl borate diester (AI-2). However, the repertoire of autoinducers produced by *Pss* B728a appears to be limited to AHLs. *N*-AHL synthase AhlI and the regulator AhlR were identified in *Pss* B728a (22) (AhlI-AhlR, Psyr1621–1622), and the single predominant AHL produced by this strain has been shown to comigrate on a chromatogram with a *N*-3-oxo-hexanoyl-L-homoserine lactone standard (23). Another potential AHL synthase similar to acyltransferases of the PlsC family is encoded by Psyr0009, which has 82% sequence identity with the HdtS protein of *P. fluorescens* F113 (24). HdtS is capable of producing three AHLs, including *N*-(3-hydroxy-7-*cis*-tetradecenoyl) homoserine lactone, a molecule also known as the *Rhizobium leguminosarum* small bacteriocin as a consequence of its growth-inhibiting properties. Homologs of both AHL synthases are present in the *Pst* DC3000 genome.

Two S2-type pyocins (Psyr0309–0310 and 4651–4652) were found in the *Pss* B728a genome, only one of which (Psyr0310) has an ortholog in *Pst* DC3000. S2 pyocins produced by *P. aeruginosa* cause death of susceptible bacteria by DNA breakdown due to endonuclease activity; in addition, they are able to inhibit phospholipid synthesis under iron-limited conditions (25). Pyocin expression in *P. aeruginosa* can be induced by treatments that cause DNA damage, but it is independent of the LexA repressor (26). However, one of the S-type pyocins in *Pss* B728a is predicted to be a member of the LexA regulon (Table 7).

Copper and antibiotic resistance. Many *P. syringae* strains are resistant to copper and streptomycin, attributable to the wide use of copper- and streptomycin-based bactericides (27). Genes responsible for both copper and streptomycin resistance are present in the genome of *Pss* B728a. Copper resistance genes, including the *copABCD* operon and a *copRS* two-component regulatory system (Psyr1493–1498), reside in a genomic island at tRNA-Lys; these proteins are 92–96% identical to the plasmid-encoded CopABCDRS proteins found in other strains of *P. syringae* (28). A copper-exporting ATPase and associated transcriptional regulator (Psyr0653–0654) might also contribute to the copper resistance of *Pss* B728a. This locus is similar to the *cueAR* system found in *P. putida* (29). Although this locus is present in *Pst* DC3000, orthologs of *copABCDRS* genes are absent, possibly explaining the lower copper tolerance of *Pst* DC3000 when compared with *Pss* B728a (S.E.L., unpublished data).

A streptomycin resistance transposon, Tn5393a, which carries a

strA-strB determinant (Psyr2669–2670), is found in *Pss* B728a, but not in *Pst* DC3000. This cassette resides near the large cluster of genes coding for biosynthesis of lipodepsipeptide toxins and the siderophore achromobactin (Fig. 3). Interestingly, this cluster also includes genes for modification of lipidA with 4-amino-4-deoxy-L-arabinose (Psyr2689–2696), which confers resistance to cationic antimicrobial peptides and antibiotics, such as polymyxin. The genome of *Pst* DC3000 contains only remnants of the lipid A modification genes, indicating that they were lost by this strain.

Ice nucleation and antifreeze proteins. A distinctive feature of *Pss* B728a is the presence of an ice nucleation gene (Psyr1608), which is inserted in an exchangeable locus (Fig. 7, which is published as supporting information on the PNAS web site). This outer-membrane ice nucleation protein acts as an ice embryo, a template for the formation of ice crystals (30). *Pss* B728a also has an unlinked gene encoding an antifreeze protein (Psyr0931) homologous to AfpA of *P. putida* GR12–2 (31). Antifreeze proteins are secreted into the medium, where they inhibit the growth of external ice by adsorbing onto the ice surface and lowering the temperature at which it can grow. The putative antifreeze gene is found in an operon with two glycosyltransferase genes (Psyr0929–0930); three genes encoding components of the type I secretion system (Psyr0933–0935) are found on the opposite strand. Most likely these genes participate in glycosylation and secretion of antifreeze protein. It is possible that the ice nucleation activity of *P. syringae* strains, which are quantitatively quite variable, is modulated by the activity of such antifreeze proteins. Orthologs of ice nucleation and antifreeze proteins are absent from the genome of *Pst* DC3000.

Osmotolerance. Many bacteria respond to decreased water availability by accumulating compatible solutes, which protect enzymes and stabilize membranes. Some of the compatible solutes accumulated by Pseudomonads include betaine, ectoine, and hydroxyectoine, *N*-α-acetylglutaminylglutamine amide, mannitol, and glucosylglycerol (32). Both *Pss* B728a and *Pst* DC3000 make betaine from choline due to the presence of choline dehydrogenase and betaine-aldehyde dehydrogenase (Psyr4732 and -4733, respectively). In addition to betaine, *Pss* B728a might also be capable of producing ectoine. Ectoine is synthesized from L-aspartate-4-semialdehyde via a three-step pathway with L-2,4 diaminobutyrate and *N*-acetyl-L-2,4 diaminobutyrate as intermediates. A gene with strong similarity to ectoine synthase (Psyr0334) is present in *Pss* B728a, but no homologs of *ectB* and *ectA*, which catalyze the first and second steps of the pathway, were found nearby. None of these genes are present in *Pst* DC3000. It is possible that this "orphan" ectoine synthase represents a remnant of an *ectABC* operon that is no longer functional. Alternatively, *N*-acetyl-L-2,4-diaminobutyrate may be produced in *Pss* B728a by a nonorthologous set of enzymes, which do not reside in an operon with ectoine synthase. Indeed, four homologs of a gene encoding L-2,4-diaminobutyrate aminotransferase are present in *Pss* B728a. Although three of them most likely participate in biosynthesis of siderophores and phytotoxins, the physiological role of the fourth gene is unknown. There is no close homolog of the acetyltransferase, *ectA*, in the genome; however, it belongs to a large family of *N*-acetyltransferases. Thirty-nine representatives of this family were found in *Pss* B728a, eight of which have no ortholog in *Pst* DC3000, and it is possible that one of these proteins is capable of catalyzing acetylation of L-2,4-diaminobutyrate. This concept is supported by the presence of "orphan" ectoine synthases in the genomes of other microbes that have been sequenced. If an ectoine synthase-like enzyme is functional in *Pss* B728a, it could explain the higher osmotolerance observed in this strain in culture, as compared with *Pst* DC3000 (33), and also contribute to higher stress tolerance of *Pss* B728a on plants (2, 3).

Metabolism. *Pss* B728a shares with *Pst* DC3000 most pathways of central metabolism, such as glycolysis, gluconeogenesis, tricarboxylic acid cycle, pentose phosphate shunt, etc. Differences between the two strains in propionate, tricaproin, L-ascorbate, L-histidine, D-tartrate, erythritol, and L-lactate utilization have been reported (34). However, most of these differences should be attributed to the differences in transcriptional regulation, because the genes responsible for utilization of these substrates are present in both genomes. Unlike *Pst* DC3000, *Pss* B728a has a gene coding for L-lactate dehydrogenase (Psyr0908), which might explain the difference in L-lactate utilization.

Virulence. Because *Pss* B728a and *Pst* DC3000 differ both in host range as well as the apparent intimacy with which they interact with host plants, it is not surprising that substantial differences in known virulence determinants are found between these two pathovars. A total of 298 virulence genes conferring a myriad of different functions have been described in *Pst* DC3000 (8), and *Pss* B728a shares many of these genes. In addition, a large set of putative virulence genes are present only in one of the two strains. Pss B728a harbors unique genes such as a cellulase family protein (Psyr4600), a pectate lyase (Psyr0852), a xylanase (Psyr4508), and a type I secretion system (Psyr3075–77).

Phytotoxins. *Pss* is known to produce a number of phytotoxins, which contribute significantly to the virulence of this strain (35). *Pss* produces two classes of lipodepsipeptides: the syringopeptins (SP) and the lipodepsinonapeptides (including SR, syringostatin, or pseudomycin). Usually each strain secretes a single type of SP and one or two lipodepsinonapeptides (36). SPs are potent necrosis-inducing phytotoxins, whereas SR enhances bacterial virulence. Both toxins also display strong antibacterial and antifungal activity. *Pss* B728a is known to synthesize two SPs and SR (37) and gene clusters for both phytotoxins (38, 39) as well as a gene encoding an ABC transporter for export of both metabolites (40) are present in the genome of *Pss* B728a, but not in *Pst* DC3000. They are part of a larger cluster (Fig. 3), which also includes genes for siderophore biosynthesis and uptake, streptomycin resistance, arginine degradation, and genes for the modification of lipid A with aminoarabinose. This cluster of 180,637 bp accounts for 2.95% of the *Pss* B728a genome and appears to be an important virulence determinant of this strain.

Pss strains are also capable of producing a family of peptide derivatives called syringolins. Syringolins have no known impact on the interaction of bacteria with their host plants, but they are recognized by nonhost plants, where these peptides activate defense-related genes and induce resistance to fungal pathogens (41). Orthologs of the genes participating in biosynthesis and export of syringolin A (42) were identified in the *Pss* B728a genome (Psyr1701–1706), but not in *Pst* DC3000.

Some pathovars of *P. syringae* produce chlorosis-inducing phytotoxins, such as coronatine and phaseolotoxin. Based on the absence of chlorotic halos around lesions induced by *Pss* B728a, this strain is presumed to lack the genes for biosynthesis of chlorosis-inducing toxins. Indeed, it lacks the genes for coronatine biosynthesis found in *Pst* DC3000. However, homologs of all six genes from one of the eight phaseolotoxin biosynthesis loci, *phtE* (43, 44), were found in *Pss* B728a (Psyr2549–2555). Orthologs of these six genes are absent from *Pst* DC3000. One additional gene, annotated in GenBank as phaseolotoxin synthetase (45), is a fragment of a pyoverdine side-chain peptide synthetase IV. It is likely that the *phtE* locus participates in biosynthesis of one of the phaseolotoxin precursors and *Pss* B728a lacks other loci for biosynthesis of phaseolotoxin itself.

The *Pss* B728a genome contains a number of genes encoding for nonribosomal peptide synthases of unknown specificity. Two of them (Psyr2576 and Psyr2577) have orthologs in *Pst* DC3000 that appear to be controlled by the sigma factor HrpL (45). Another (Psyr3722) has an ortholog in Pst DC3000 that is truncated and therefore likely to be nonfunctional or to produce a different peptide. A third gene cluster includes three putative nonribosomal peptide synthases (Psyr1792–1795) with orthologs in *P. fluorescens* Pfo-1 but not in *Pst* DC3000. One of the subunits of this synthase is a hybrid nonribosomal peptide synthase/polyketide synthase. An operon encoding subunits of a type I polyketide synthase is present

in the genome of *Pss* B728a. These genes have some similarity to the *pks* operon in *Bacillus subtilis* 168, which makes a polyketide compound of unknown structure. Thus, *Pss* B728a is able to produce a wide repertoire of secondary metabolites; however, their biological activity and physiological role have yet to be determined.

Secreted factors. *Pss* B728a has 27 type III secretion system (TTSS) effectors, five of which are not found in *Pst* DC3000 (www.Pseudomonas-syringae.org/pst_func_gen2.htm). These five include the four unique effectors identified by Greenberg and Vinatzer (46), and another putative unique effector (Psyr1890). This effector bears 41% identity in its C-terminal part to the type III effector HopPtoH and 28% identity over its entire length to a non-LEE-encoded (LEE, Locus for Enterocyte Effacement) type III effector D from *Citrobacter rodentium*. We identified two additional putative TTSS effectors common to both organisms, namely *avrF1* and *shcZ3* (Psyr1187 and Psyr1225, respectively). *Pst* DC3000 has 28 more TTSS effectors than does *Pss* B728a, having a total of 55 effectors (www.Pseudomonas-syringae.org/pst_func_gen2.htm). The distinct complement of effectors present in each strain should prove helpful in deciphering the common and unique functions of these proteins.

Auxin. Bacterial indole-3-acetic acid (IAA) can have either a stimulatory or inhibitory effect on plant growth depending on its concentration and the plant tissue exposed (47). Most pathovars of *P. syringae* produce IAA (48), but the amounts produced vary greatly between strains. *Pss* B728a produces substantially higher levels of IAA than *Pst* DC3000 in most culture media (T. Powell, personal communication), and IAA production is known to contribute to virulence and epiphytic fitness of *Pss* B728a (47). *Pss* B728a has an operon for biosynthesis of IAA, which includes tryptophan monooxygenase, *iaaM* (Psyr1536) and indoleacetamide hydrolase *iaaH* (Psyr1537), and genes first described in *P. syringae* pv. *savastanoi* (49). The genes in the operon are >90% identical to those in the IAA operon of *Pss* Y30 (47). Although *Pst* DC3000 has no orthologs of *iaaM* and *iaaH*, it produces IAA (8), possibly due to PSPTO0517 and PSPTO0518, which have low sequence similarity to tryptophan monooxygenase and indoleacetamide hydrolase, respectively. Orthologs of these genes are also found in *Pss* B728a.

Pss B728a might additionally produce IAA via a pathway in which indole-3-acetaldoxime, not tryptophan, is the initial precursor. Indole-3-acetaldoxime is a common intermediate in two plant pathways of IAA biosynthesis, and in cruciferous plants it also serves as a branching point in biosynthesis of IAA (49), indole phytoalexins (such as camalexin), and indole glucosinolates (50). Both phytoalexins and glucosinolates are involved in plant defense. A chromosomal cluster found in *Pss* B728a includes a homolog of *Rhodococcus globerulus* aldoxime dehydratase (Psyr0006), which is active on both aliphatic and arylaliphatic aldoximes and a nitrilase (Psyr0007). These two enzymes would catalyze conversion of indole-3-acetaldoxime into indole-3-acetonitrile and IAA, respectively, thus redirecting the flux of indole-3-acetaldoxime toward IAA rather than phytoalexins and glucosinolates (provided that the plant host produces any of those). This would circumvent one of the plant defense mechanisms and might even inhibit the plant hypersensitive response (51). An ortholog of nitrilase is present in *Pst* DC3000, but an aldoxime dehydratase is apparently a pseudogene disrupted by an IS element. Further metabolism of auxin also appears to differ between *Pss* B728a and *Pst* DC3000. A gene encoding indoleacetate–lysine ligase, which metabolizes IAA to an amino acid conjugate devoid of auxin activity, is present in *Pst* DC3000 (PSPTO371), but not in *Pss* B728a.

Conclusion

Pss B728a and *Pst* DC3000 differ markedly from one another with respect to host range and manner of association with host plants. A large number of gene differences that likely contribute to such differential behavior have been described here. The challenge remains to associate these genes to specific traits of the bacterial pathogens. As more strains of *P. syringae* are sequenced, comparative genomic analyses should prove useful in deciphering virulence and behavioral traits of this important species of phytopathogenic bacteria.

We thank the members of the Joint Genome Institute who contributed to this work. Funding for this work described in this study was provided by the Department of Energy, Office of Biological and Environmental Research.

1. Sawada, H., Suzuki, F., Matsuda, I. & Saitou, N. (1999) *J. Mol. Evol.* **49,** 627–644.
2. Hirano, S. S. & Upper, C. D. (2000) *Microbiol. Mol. Biol. Rev.* **64,** 624–653.
3. Boureau, T., Routtu, J., Roine, E., Taira, S. & Romantschuk, M. (2002) *Mol. Plant Pathol.* **3,** 451–460.
4. Beattie, G. A. & Lindow, S. E. (1999) *Phytopathology* **89,** 353–359.
5. Loper, J. E. & Lindow, S. E. (1987) *Phytopathology* **77,** 1449–1454.
6. Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., *et al.* (2003) *J. Bacteriol.* **185,** 2759–2773.
7. Aranda-Olmedo, I., Tobes, R., Manzanera, M., Ramos, J. L. & Marques, S. (2002) *Nucleic Acids Res.* **30,** 1826–1833.
8. Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10181–10186.
9. Jensen, L. J., Skovgaard, M., Sicheritz-Ponten, T., Hansen, N. T., Johansson, H., Jorgensen, M. K., Kiil, K., Hallin, P. F. & Ussery, D. (2004) in *Pseudomonas, Life Style and Molecular Architecture*, ed. Ramos, J. L. (Kluwer Academic, New York), Vol. 1, p. 158.
10. Heidelberg, J. F., Paulsen, I. T., Nelson, K. E., Gaidos, E. J., Nelson, W. C., Read, T. D., Eisen, J. A., Seshadri, R., Ward, N., Methe, B., *et al.* (2002) *Nat. Biotechnol.* **20,** 1118–1123.
11. Klockgether, J., Reva, O., Larbig, K. & Tummler, B. (2004) *J. Bacteriol.* **186,** 518–534.
12. Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S. A., Mavrodi, D. V., DeBoy, R. T., Seshadri, R., Ren, Q., Madupu, R., *et al.* (2005) *Nat. Biotechnol.* **23,** 873–878.
13. Meyer, J. M. (2000) *Arch. Microbiol.* **174,** 135–142.
14. Ravel, J. & Cornelis, P. (2003) *Trends Microbiol.* **11,** 195–200.
15. Franza, T., Mahé, B. & Expert, D. (2005) *Mol. Microbiol.* **55,** 261–275.
16. Muzinger, M., Budzikiewicz, H., Expert, D., Enard, C. & Meyer, J. M. (2000) *Naturforschung* **55,** 328–332.
17. Bell, K. S., Sebaihia, M., Pritchard, L., Holden, M. T. G., Hyman, L. J., Holeva, M. C., Thomson, N. R., Bentley, S. D., Churcher, L. J. C., Mungall, K., *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11105–11110.
18. Braun, V., Mahren, S. & Ogierman, M. (2003) *Curr. Opin. Microbiol.* **6,** 173–180.
19. Jacobs, J. L., Carroll, T. L. & Sundin, G. W. (2005) *Microbial Ecol.* **49,** 104–113.
20. Erill, I., Escribano, M., Campoy, S. & Barbe, J. (2003) *Bioinformatics* **19,** 2225–2236.
21. Sundin, G. W., Jacobs, J. L. & Murillo, J. (2000) *Appl. Environ. Microbiol.* **66,** 5167–5173.
22. Quinones, B., Pujol, C. J. & Lindow, S. E. (2004) *Mol. Plant–Microbe Interact.* **17,** 521–531.
23. Cha, C., Gao, P., Chen, Y.-C., Shaw, P. D. & Farrand, S. K. (1998) *Mol. Plant–Microbe Interact.* **11,** 1119–1129.
24. Laue, B. E., Jiang, Y., Chhabra, S. R., Jacob, S., Stewart, G. S., Hardman, A., Downie, J. A., O'Gara, F. & Williams, P. (2000) *Microbiology* **146,** 2469–2480.
25. Michel-Briand, Y. & Baysse, C. (2002) *Biochimie* **84,** 499–510.
26. Matsui, H., Sano, Y., Ishihara, H. & Shinomiya, T. (1993) *J. Bacteriol.* **175,** 1257–1263.
27. Sundin, G. W. & Bender, C. L. (1993) *Appl. Environ. Microbiol.* **59,** 1018–1024.
28. Mill, S. D., Jasalavich C. A. & Cooksey D. A. (1993) *J. Bacteriol.* **175,** 1656–1664.
29. Adaikkalam, C. & Swarup, S. (2002) *Microbiology* **148,** 2857–2867.
30. Green, R. L. & Warren. G. J. (1985) *Nature* **317,** 645–648.
31. Muryoi, N., Sato, M., Kaneko, S., Kawahara, H., Obata, H., Yaish, M. W. F., Griffith, M. & Glick, B. R. (2004) *J. Bacteriol.* **187,** 5661–5671.
32. Kets, E. P. W., Galinski, E. A., de Wit, M., de Bont, J. A. M. & Heipieper, H. J. (1996) *J. Bacteriol.* **178,** 6665–6670.
33. Wright, C. A. & Beattie, G. A. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 3269–3274.
34. Sands, D. C., Schroth, M. N. & Hildebrand, D. C. (1970) *J. Bacteriol.* **101,** 9–23.
35. Bender, C. L., Alarcon-Chaidez, F. & Gross, D. C. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 266–292.
36. Lu, S-E., Soule, J. D. & Gross, D. C. (2003) *Appl. Environ. Microbiol.* **69,** 7273–7280.
37. Grgurina, I., Mariotti, F., Fogliano, V., Gallo, M., Scaloni, A., Iacobellis, N. S., Lo Cantore, P., Mannina, L., van Axel Castelli, V., Greco, M. L., *et al.* (2002) *Biochim. Biophys. Acta* **1597,** 81–89.
38. Scholz-Schroeder, B. K., Hutchison, M. L., Grgurina, I. & Gross, D. C. (2001) *Mol. Plant–Microbe Interact.* **14,** 336–348.
39. Guenzi E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. (1998) *J. Biol. Chem.* **273,** 32857–32863.
40. Scholz-Schroeder, B. K., Soule, J. D., Lu, S. E., Grgurina, I. & Gross, D. C. (2001) *Mol. Plant–Microbe Interact.* **14,** 1426–1435.
41. Waspi, U., Schweizer, P. & Dudler, R. (2001) *Plant Cell* **13,** 153–161.
42. Amrein, H., Makart, S., Granado, J., Shakaya, R., Schneider-Pokorny, J. & Dudler, R. (2004) *Mol. Plant–Microbe Interact.* **17,** 90–97.
43. Zhang, Y., Rowley, K. B. & Patil, S. S. (1993) *J. Bacteriol.* **175,** 6451–6458.
44. Zhang, Y. X. & Patil, S. S. (1997) *Mol. Plant–Microbe Interact.* **10,** 947–960.
45. Fouts, D. E., Abramovitch, R. B., Alfano, J. R., Baldo, A. M., Buell, C. R., Cartinhour, S., Chatterjee, A. K., D'Ascenzo, M., Gwinn, M. L., Lazarowitz, S. G., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 2275–2280.
46. Greenberg, J. T. & Vinatzer, B. A. (2003) *Curr. Opin. Microbiol.* **6,** 20–28.
47. Mazzola, M. & White, F. F. (1994) *J. Bacteriol.* **176,** 1374–1382.
48. Glickmann, E., Gardan, L., Jacquet, S., Hussain, S., Elasri, M., Petit, A. & Dessaux, Y. (1998) *Mol. Plant–Microbe Interact.* **11,** 156–162.
49. Cohen, J. D., Sloven, J. P. & Hendrickson, A. M. (2003) *Trends Plant Sci.* **8,** 197–199.
50. Glawischnig, E., Hansen, B. G., Olsen, C. E. & Halkier, B. A. (2004) *Proc. Natl. Acad. Sci. USA.* **101,** 8245–8250.
51. Robinette, D. & Matthysse, A. G. (1999) *J. Bacteriol.* **172,** 5742–5749.